Lending Club Case Study

Anirudh Grover
Dr. Pandur Venkatesan Balaji

EPGP-IIITB-UPGRAD-ML C54

Abstract

About Company:

Lending Club, a consumer finance company which specializes in lending various types of loans to urban customers. Lending Club is the largest online loan marketplace facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Business Understanding :

When Lending Company receives a loan application, the company has to make a decision for loan approval based on the applicant's
profile. Two types of risks are associated with the bank's decision:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders.

Business Objective :

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Goals of Data Analysis

Goals:

1. Mitigate high risky lending.
2. Reduce credit loss for company by identifying risky loan applicants.

Objectives:

- Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders.
- The objective of this EDA is to identify the risky loan applicants at the time of loan application. So that such loans can be reduced and thereby cutting down the amount of credit loss
- In other words, to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- The company can utilize this knowledge for its portfolio and risk assessment. And thus minimize the risk of losing money while lending to customers.

Problem Solving Methodology

Data Understanding
- Business Objective
- Understanding Metadata/Data Dictionary

Data Cleaning
- Fixing Rows & Columns
- Treating Missing & Invalid Values
- Standardizing Values
- Filtering Data

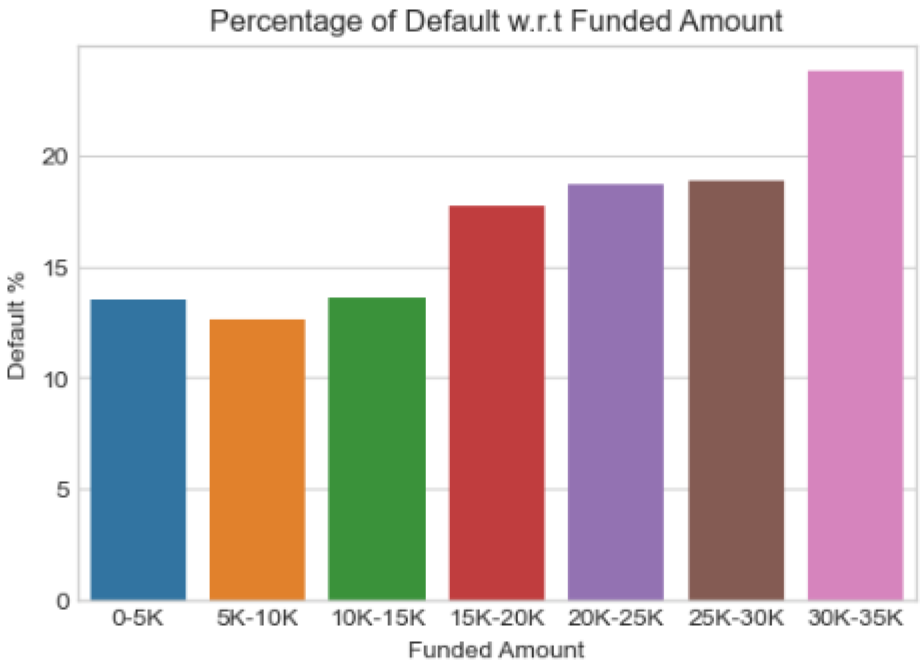
Univariate Analysis
- Analysis of Each Variable
- Plotting Bar Plots To Gain Useful Insights

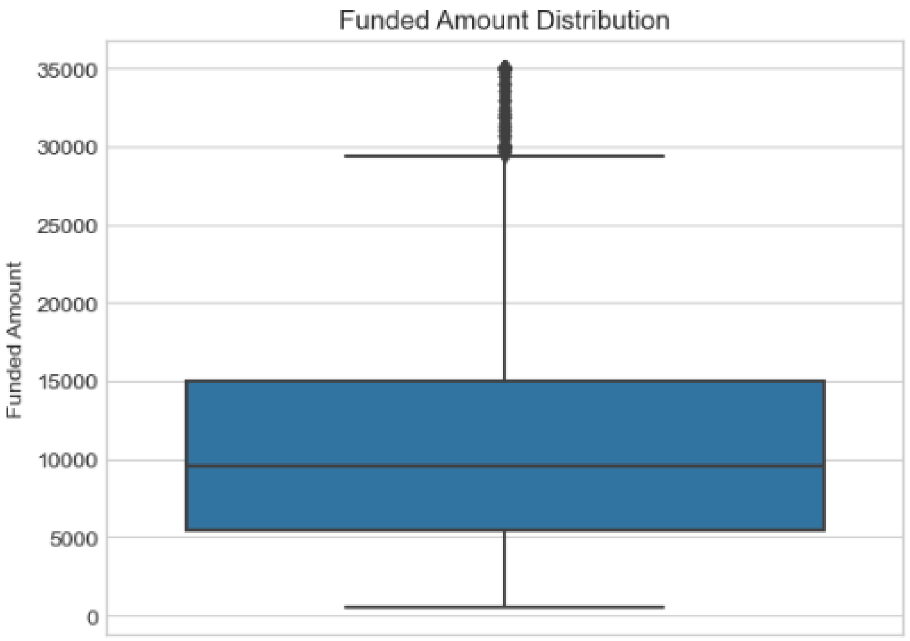
Univariate Segmented Analysis
- Grouping Variables To Get More Insights
- Plotting Bar Plots

Bivariate Analysis
- Getting Detailed Insights About Variables Obtained From Univariate Analysis To Identify Pairs of Related Variables.

Conclusion

Univariate Analysis – Funded Amount





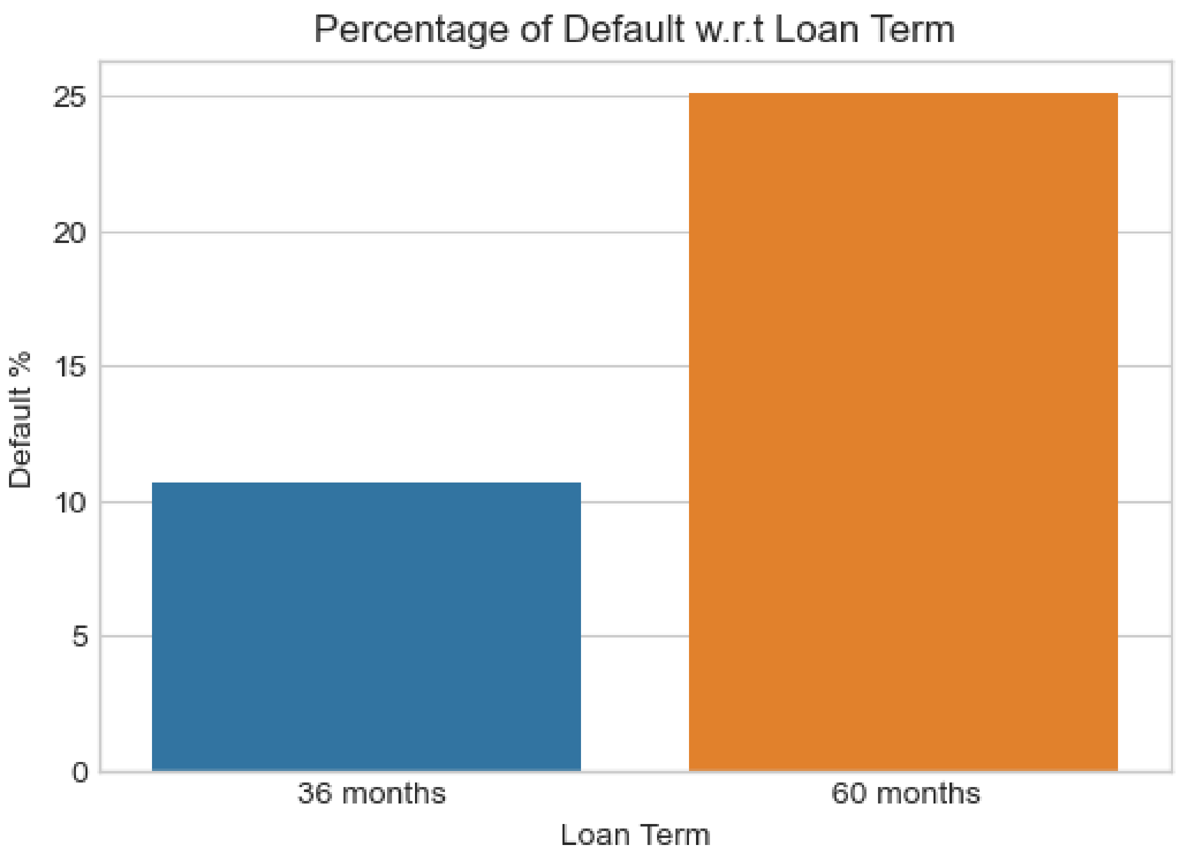
Higher Funding Amounts leads to High-Interest Amounts which becomes difficult for the applicant to cover, which results in a higher Default Rate.

Funding Amount in range of $30K-$35K has around 24% Default Rate, much higher than other categories.

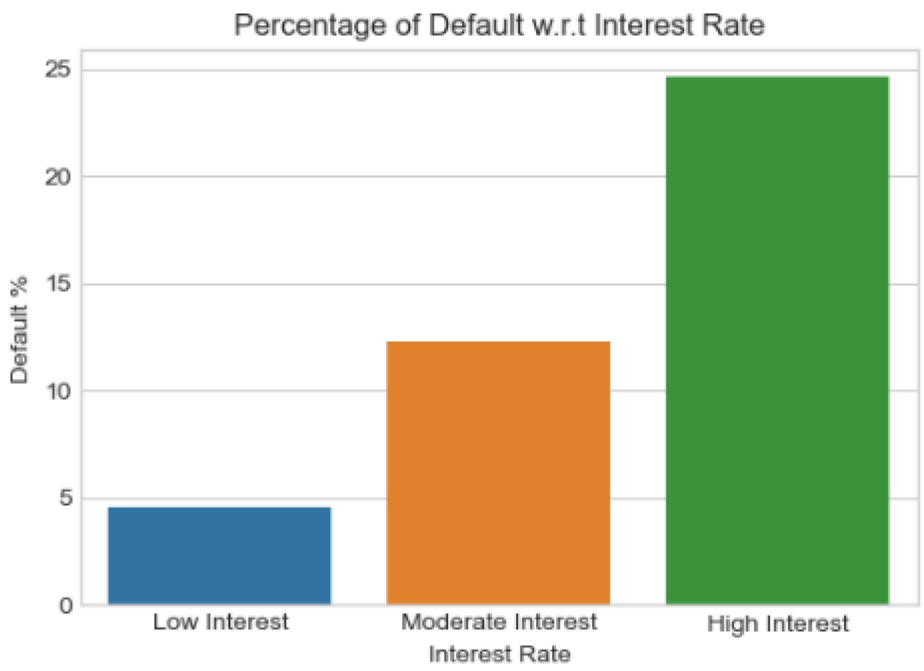
Univariate Analysis – Loan Term

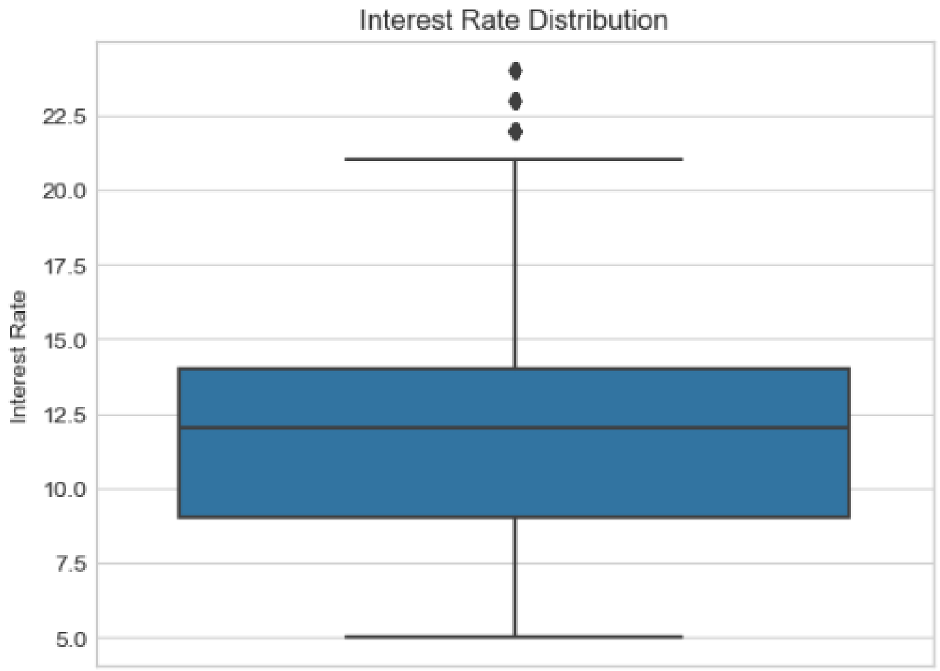
- **The loans with 60 months term tends to be defaulted more in comparison to loans taken for 36 months.**

- The reason for this can be that the applicants taking a loan for a longer term will have a higher interest amount and hence the chances of them defaulting on the loan is more



Univariate Analysis – Interest Rate





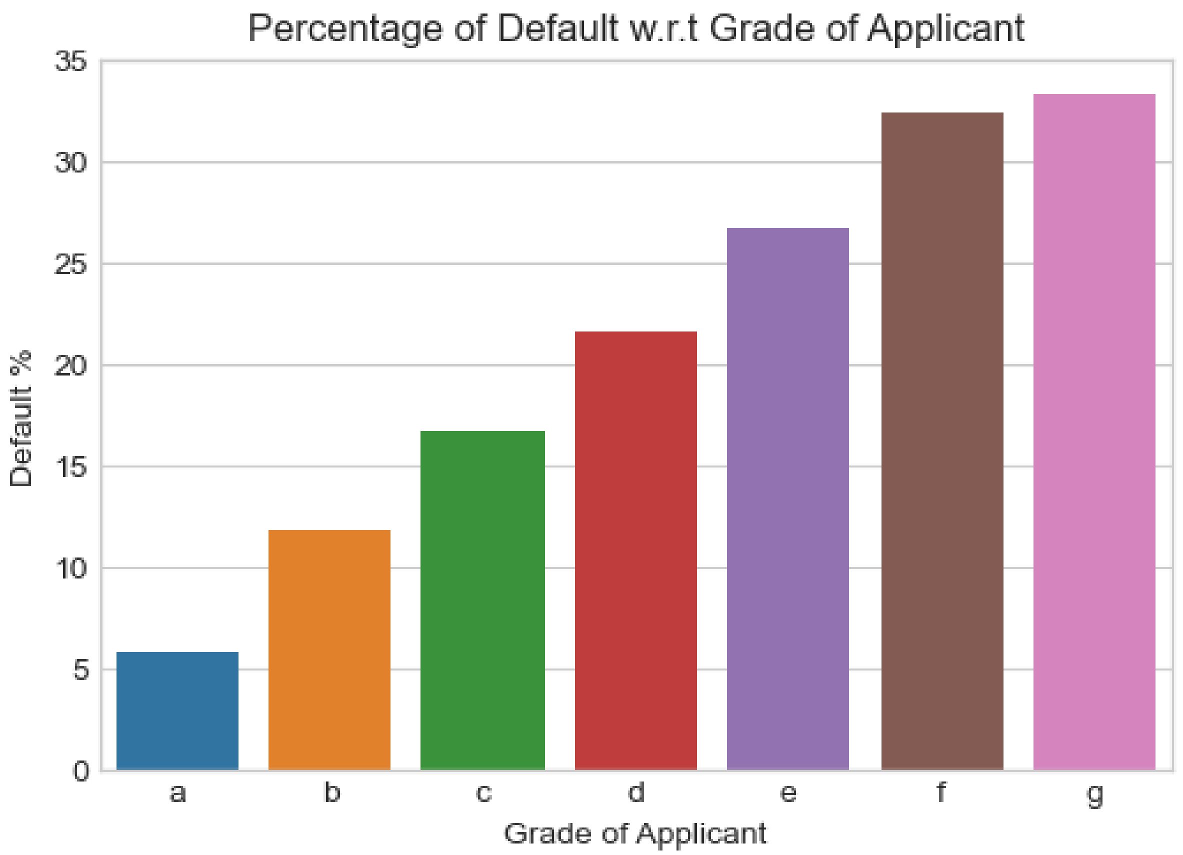
The higher the Interest for the loan, the more likely applicant will default on their loan.

Loans with Higher Interest Rate will have a Higher Interest Amount. Hence, the chances of them defaulting on the loan are more.

Univariate Analysis – Grade

- **Lower the grade, more likely applicant will default on their loan.**

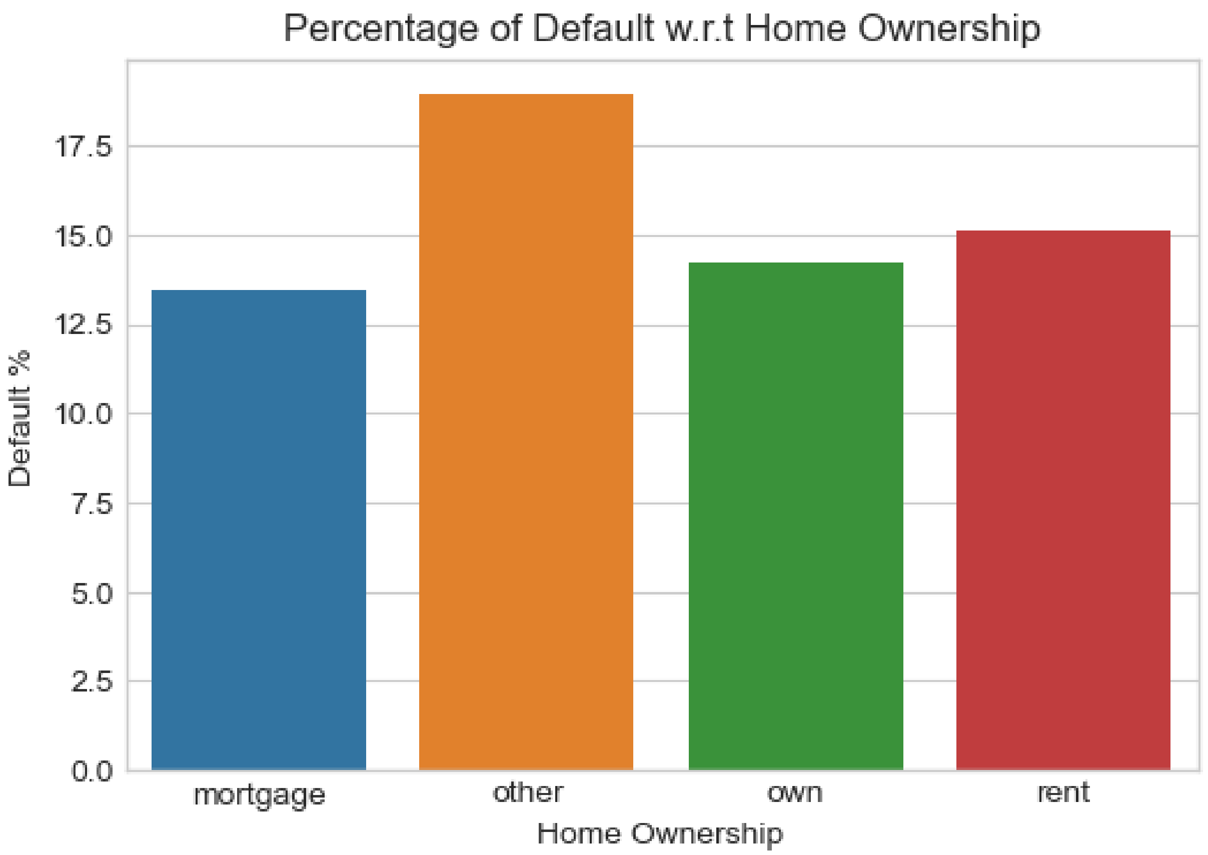
- As lower grade means that the applicant credit history is not good and hence he/she is more likely to default



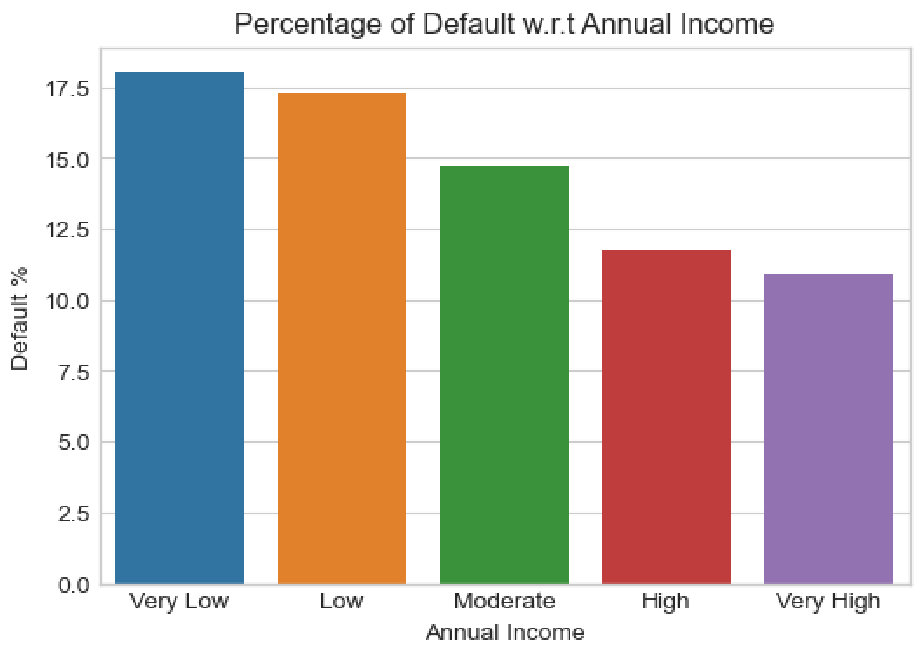
Univariate Analysis – Home Ownership Type

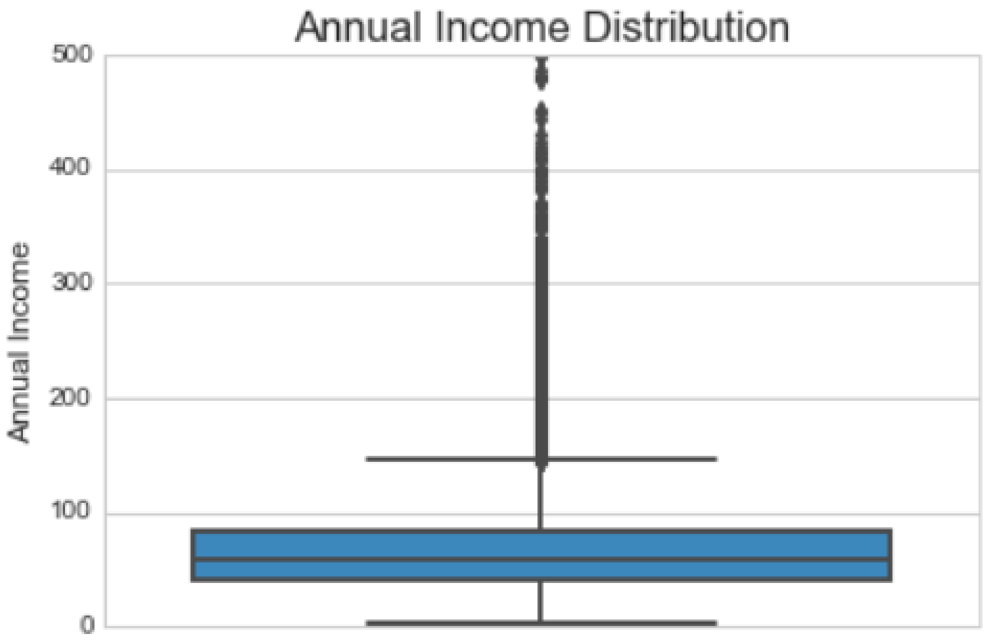
- **The applicants with *other* type of home ownership are most likely to default on their loans. If we ignore the other type, we are unable to find a strong relationship between type of home ownership and default rate.**

- The reason behind this is that type of home ownership can be supposed to the economic status of the applicant.



Univariate Analysis – Annual Income



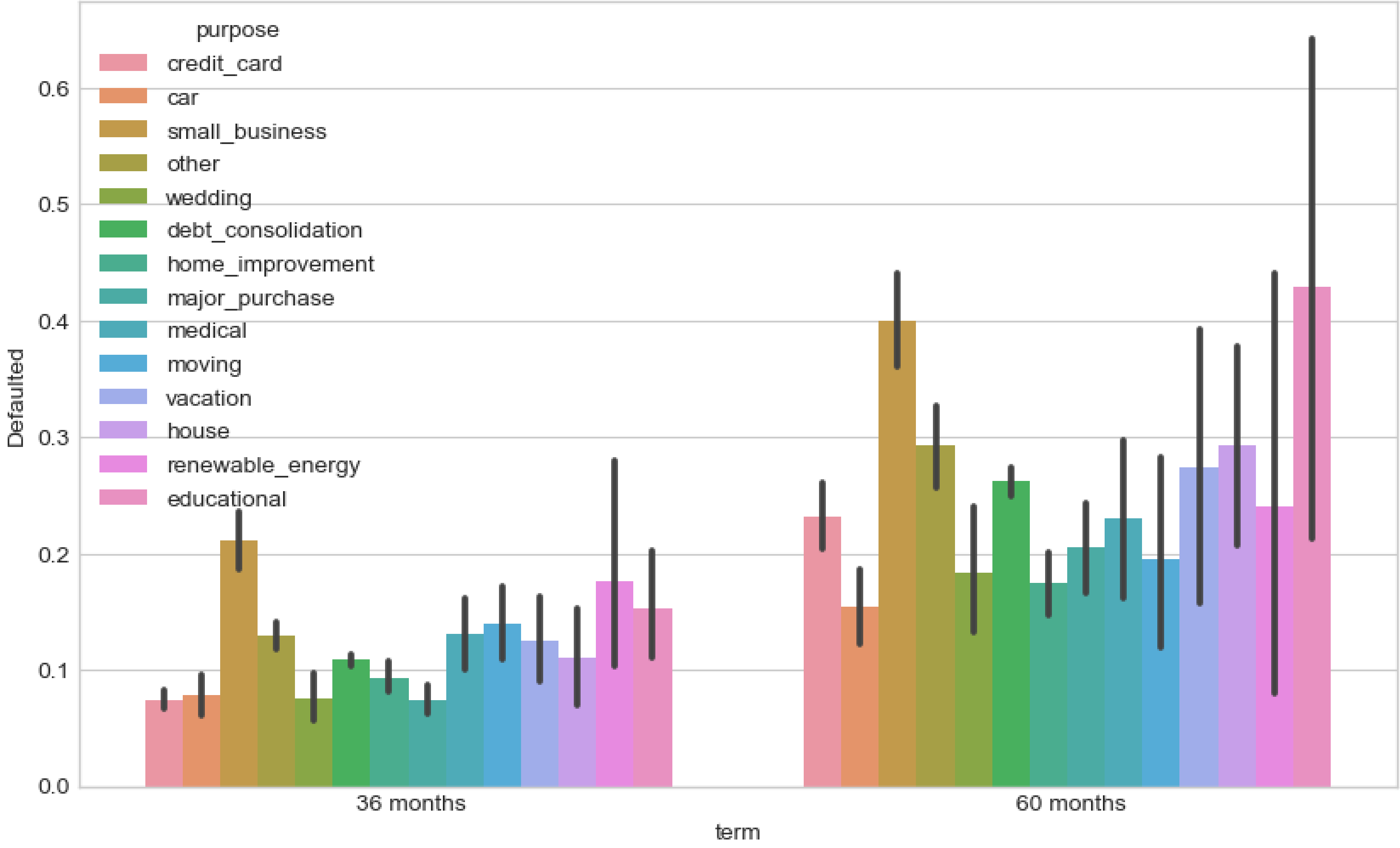


- **The applicants with annual income lying in either low or very-low category are the one who are most like to default.**

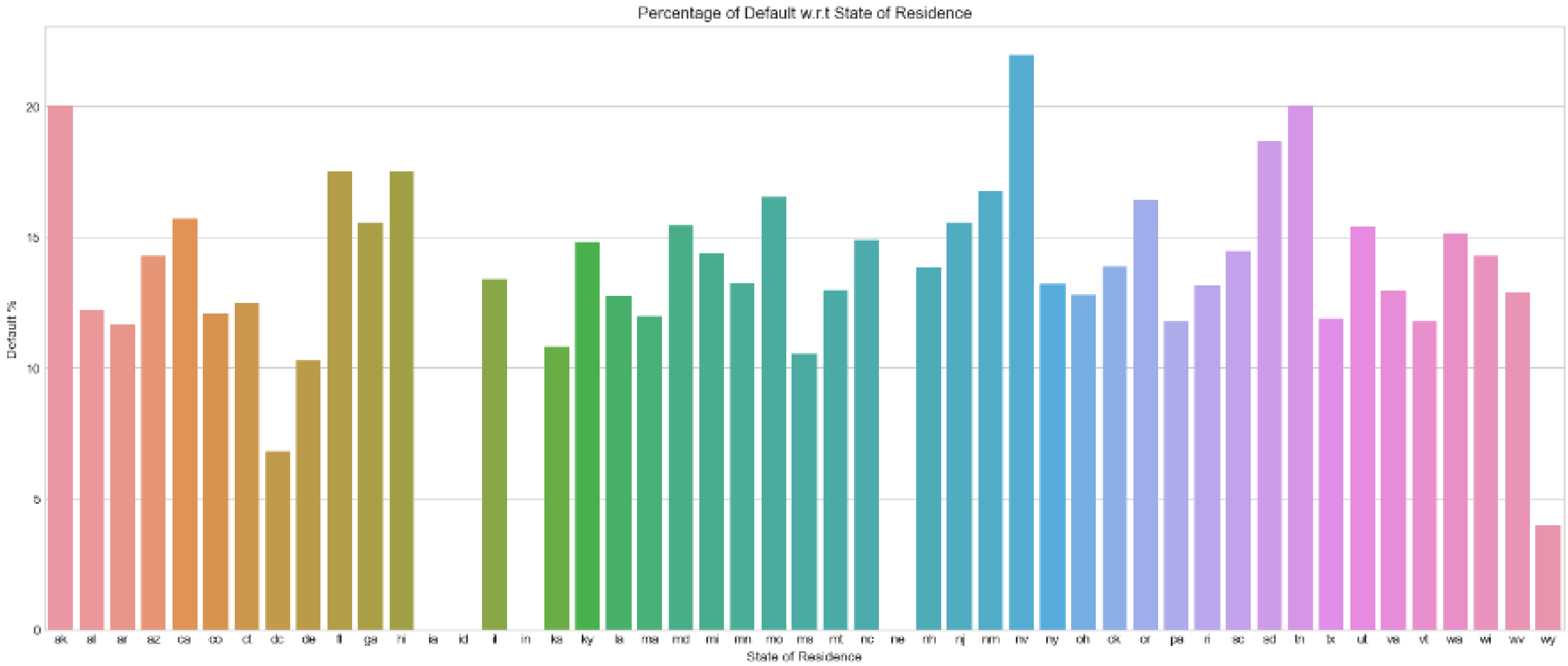
- Lower the income, less likely to cover up the loan for an applicant as they might face some financial issues. Applicants having income less than $40,000 are the ones who are more likely to default.

Univariate Analysis – Loan Purpose

- **In both the terms the lenders with the purpose of small-business, others, renewable energy, and educational are more probable defaulters**



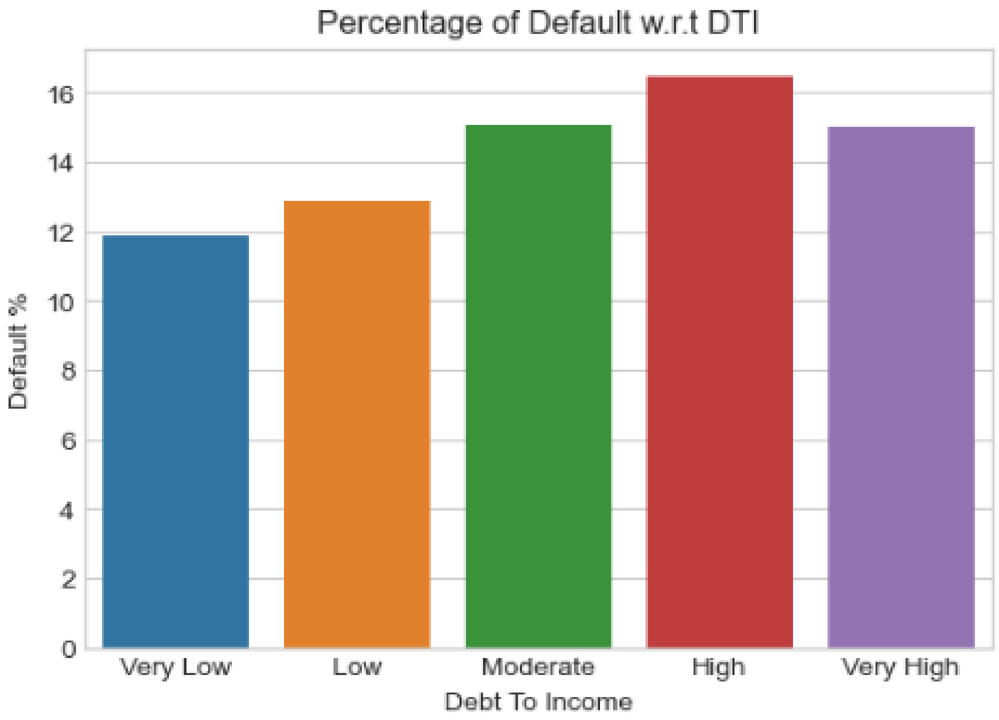
Univariate Analysis – Annual Income



Applicants belonging to State NV(Nevada), TN(Tenessee), and AK(Alaska) are most likely to default.

People belonging to IA(Iowa), ID(Idaho), NE(Nebraska), and IN(Indiana) rarely defaults on their loans.

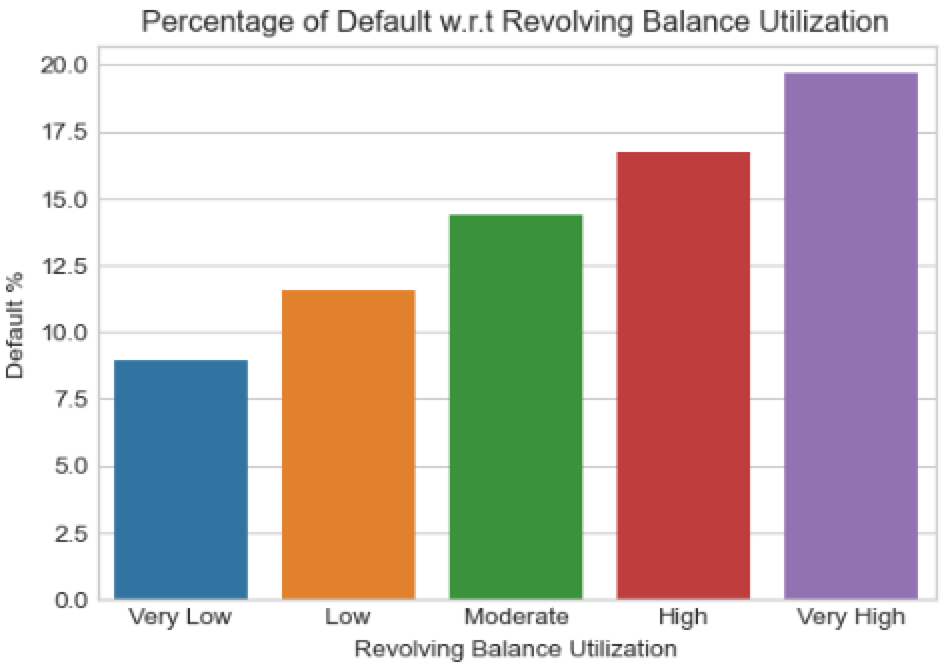
Univariate Analysis – Debt To Income (DTI)



As DTI increases, the chances of default also increases.

The reason for this is DTI stands for Debt To Income Ratio. If DTI is high, the financial condition of the person is most probably poor and hence they are more likely to default.

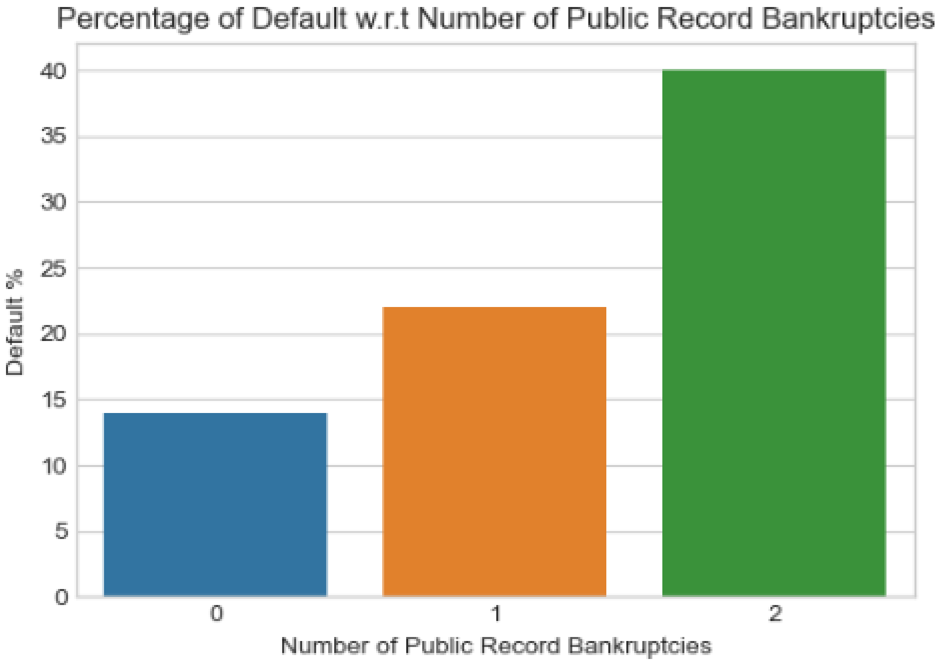
Univariate Analysis – Revolving Balance Utilization



More is the Revolving Balance Utilization more is the tendency to Default.

Higher Revolving Balance Utilization means the person has not much money to spend. That's why he/she is more likely using credit card balance and is more reliant on credits/revolving balance which in turn results in Default as it's a very risky approach to completely depend upon credit balance.

Univariate Analysis – Public Record Bankruptcies

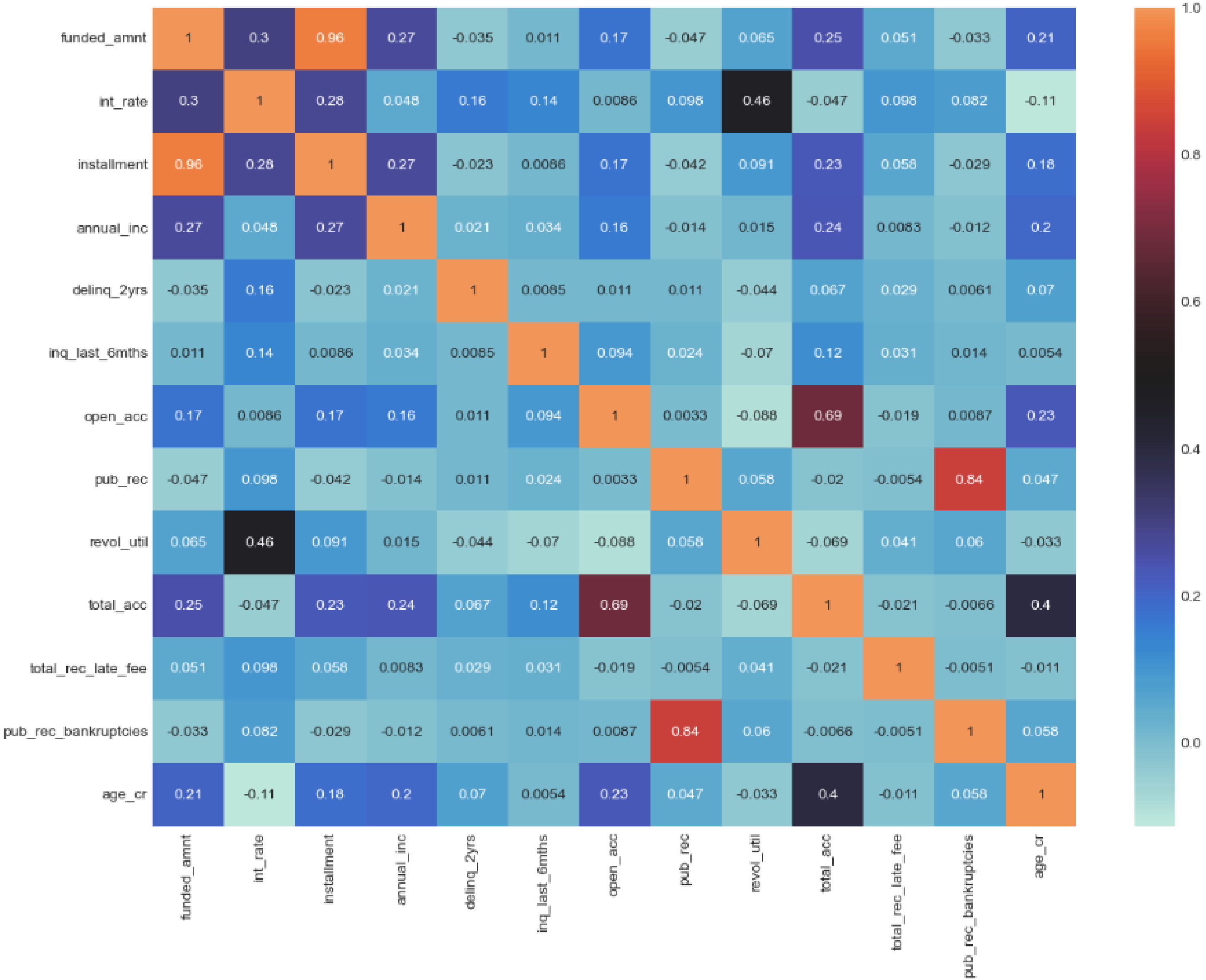


Applicants with a high number of Bankruptcies tend to default on their loans i.e. in the range of 20-40%.

Applicants who have already been are in poor financial conditions even after recovering from bankruptcy. This increases their chances of defaulting on any loan.

The number of Bankruptcy plays a major role in leading to Default.

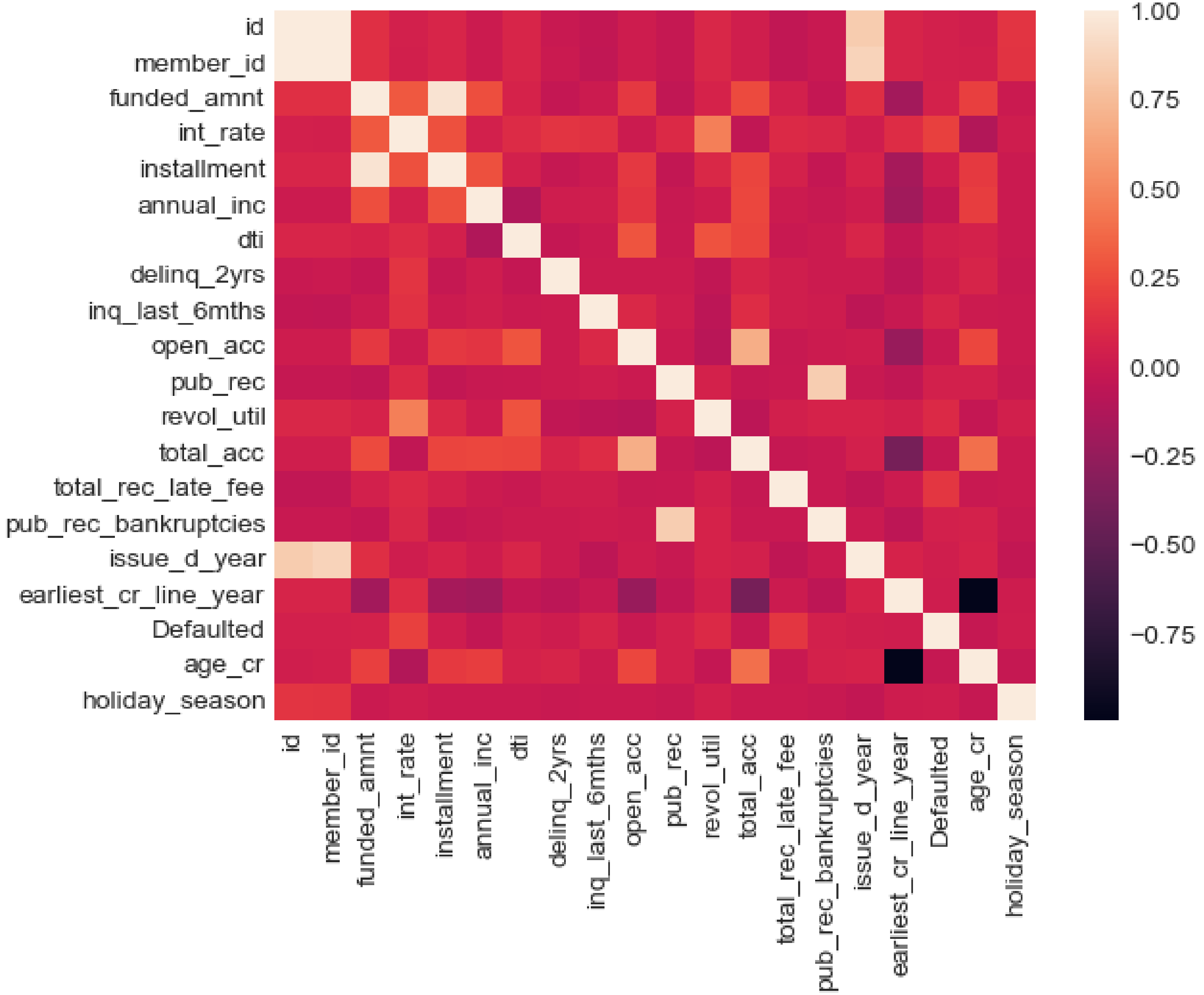
Bivariate Analysis – Correlation



- Funded Amount and Installment are very highly correlated (0.96)
- Public Derogatory Records and Public Record Bankruptcies are very highly correlated (0.84)
- Open Credit Lines and Total Credit Lines are highly correlated(0.69)
- Interest Rate and Revolving Balance Utilization are moderately correlated (0.46)
- Total Credit Lines and Age of Credit History are moderately correlated(0.40)

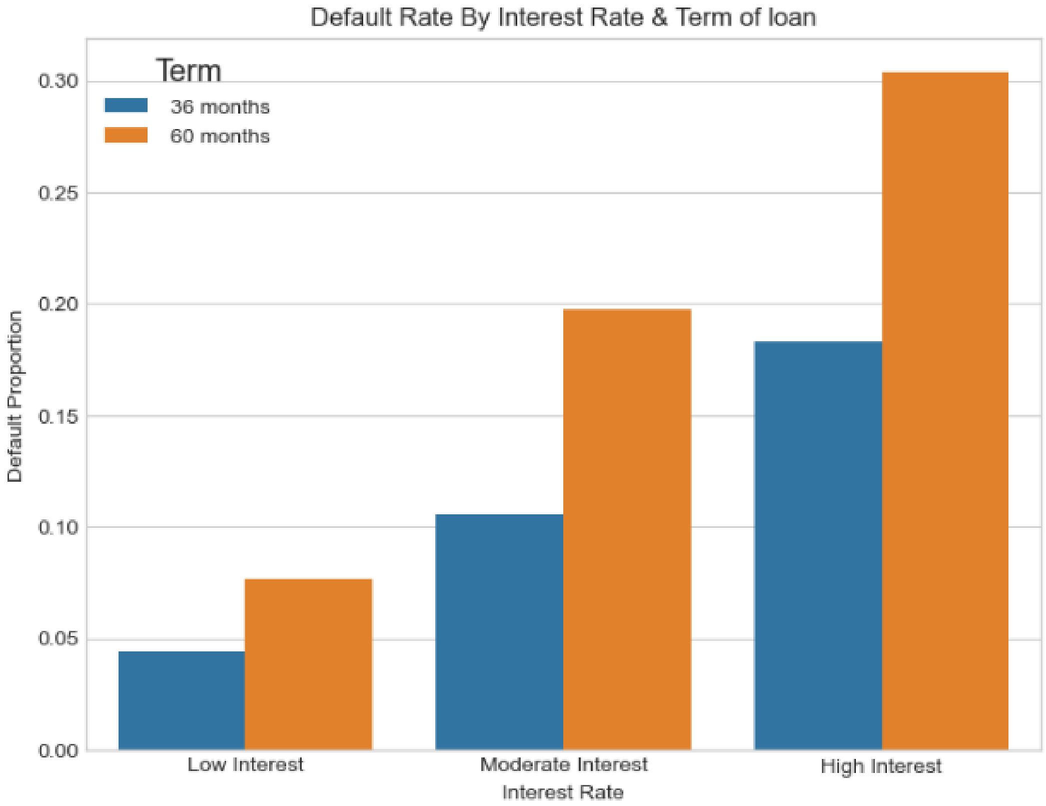
Full Correlation

- Among all the variables the interest rate tend to affect the probability of defaulter the most



Bivariate Analysis

Default Rate By Interest Rate & Term of Loan

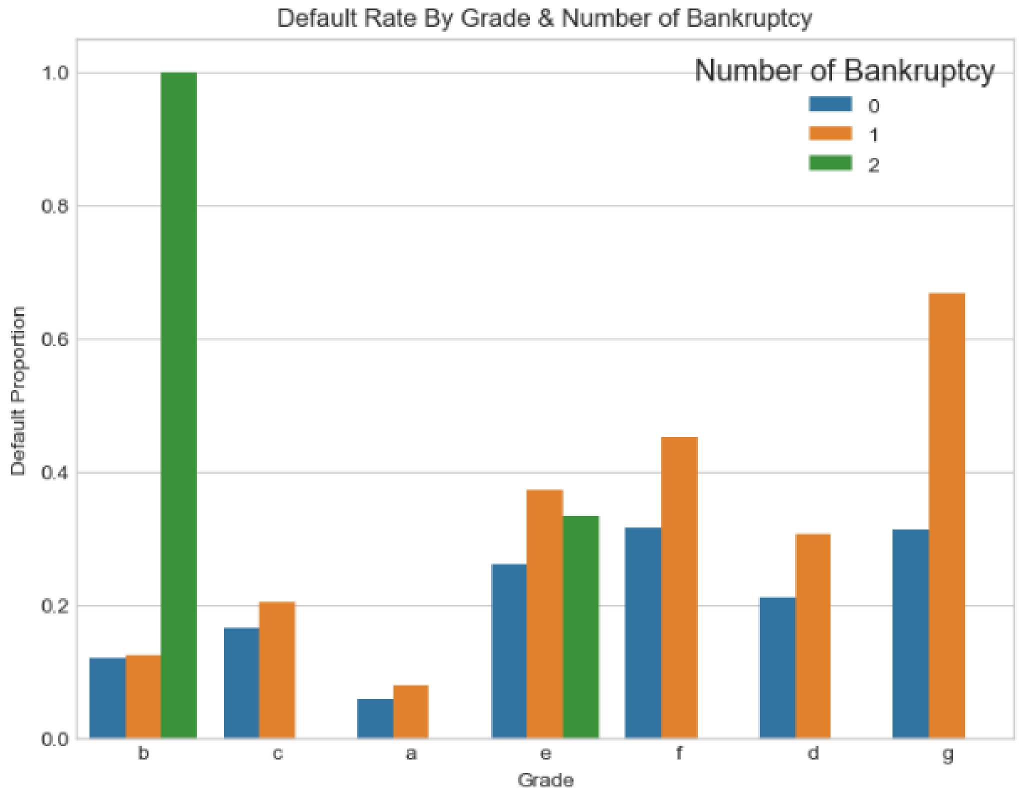


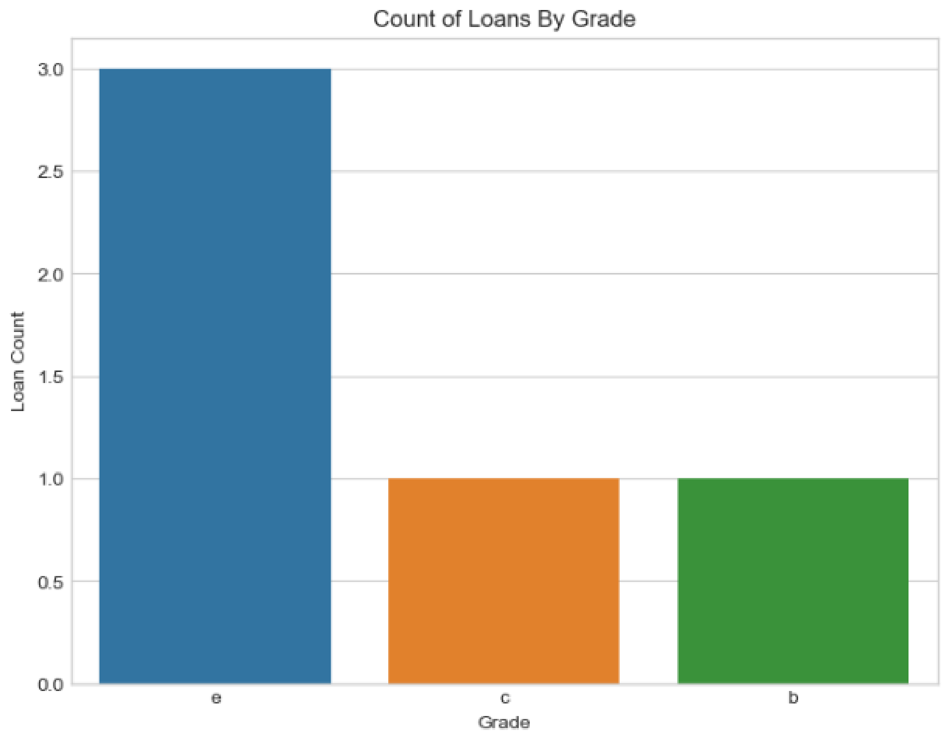
- **For Interest Rate, Loans with High-Interest Rate is most likely to default in every category of the term.**
- **For Term, Longer-term of Loan is having a high default rate for all of the Interest Rate types.**

Combination of High-Interest Rate with Longer Loan Term is an important combination for determining default rate.

Bivariate Analysis

Default Rate By Grade & Number of Bankruptcy



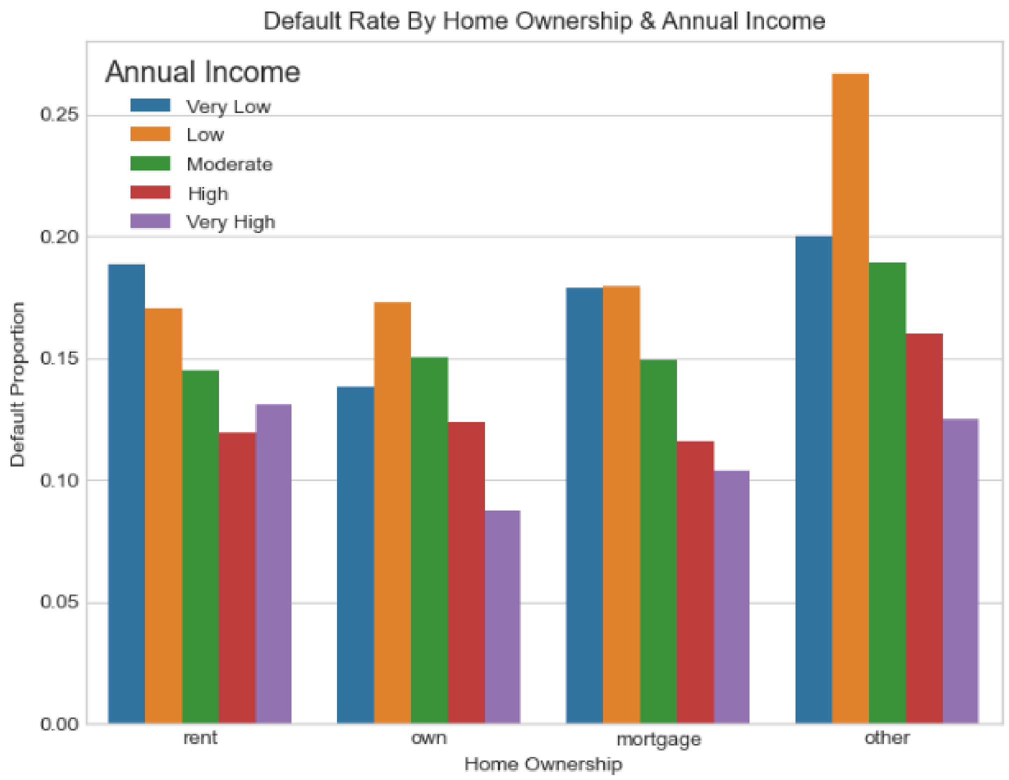


- **For Grade, grades D, E, F, G are most likely to default in every category of Number of Bankruptcy.**
- **For the Number of Bankruptcy, applicants with bankruptcy are having a high default rate for all of the grade types.**

The combination of Low grades i.e. D, E, F, G with Bankruptcy i.e. 1+ is an important combination for determining default rate.

Bivariate Analysis

Default Rate By Home Ownership & Annual Income

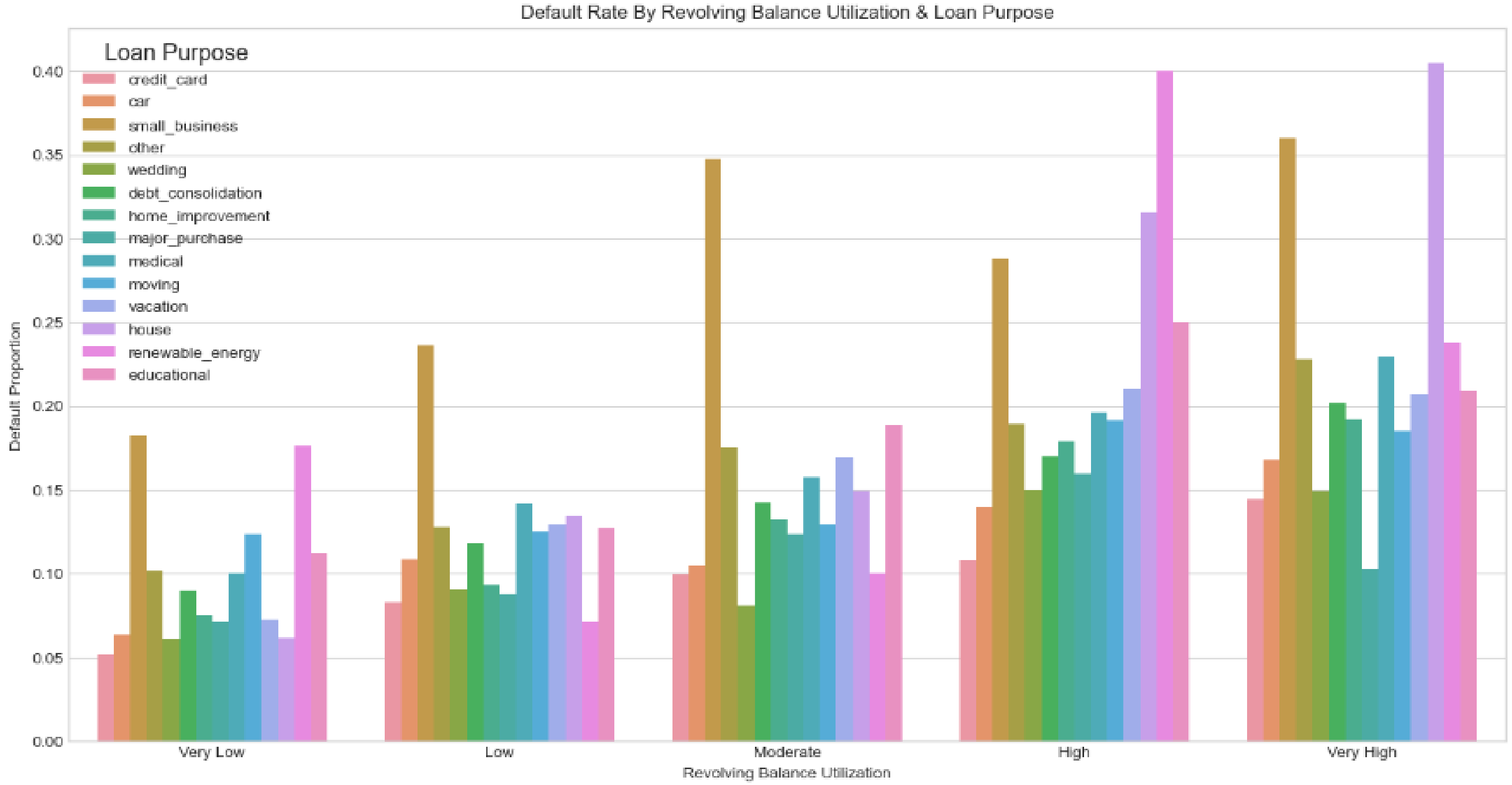


- **For Home Ownership, the “Other” ownership type is most likely to default in every category of income.**
- **For Annual Income, applicants with Low income are having a high default rate for all of the home ownership types.**

The combination of home ownership type as Other and Low annual income is an important combination for determining default rate.

Bivariate Analysis

Default Rate By Revolving Balance Utilization & Loan Purpose

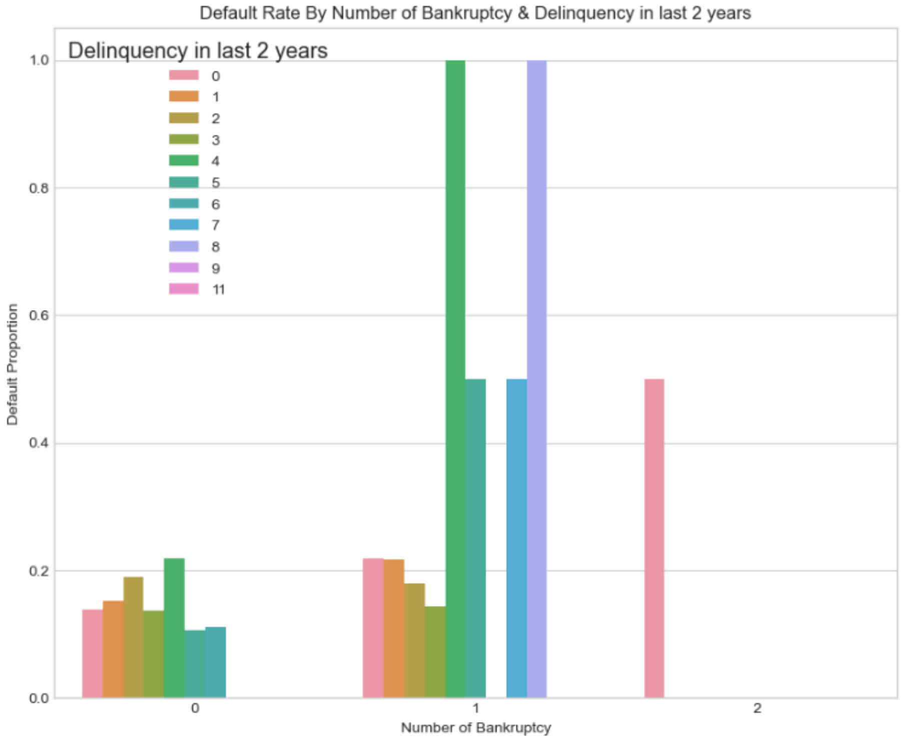


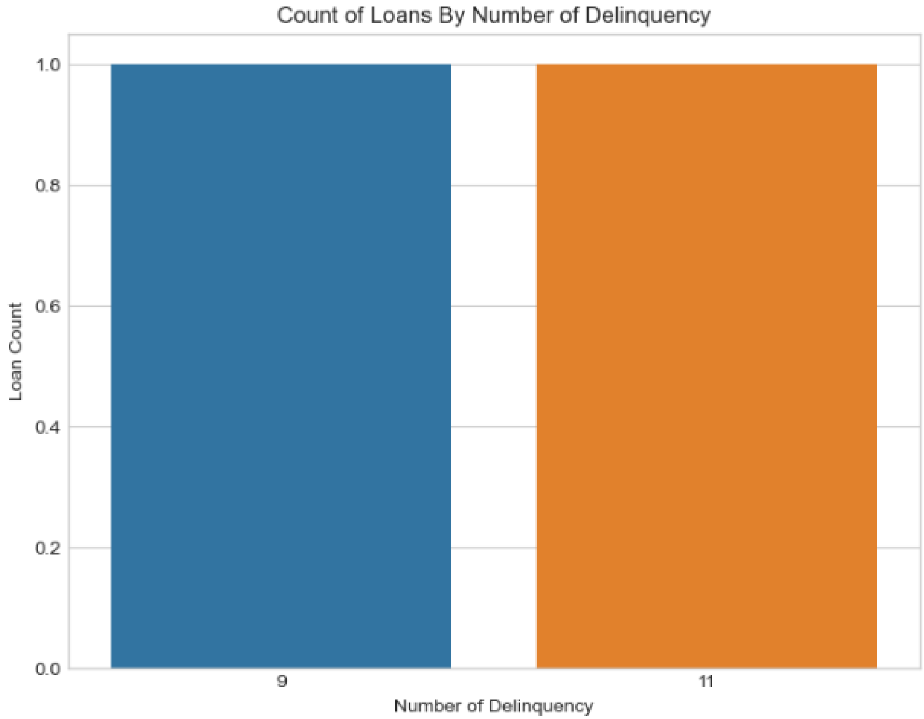
- **For Revolving Balance Utilization, Very High revolving balance utilization is most likely to default in every category of loan purpose.**
- **For Loan Purpose, applicants with Small Business as loan purpose are having high default rate for all of the revolving balance utilization categories.**

The combination of Very High revolving balance utilization and Small Business as loan purpose is an important combination for determining default rate.

Bivariate Analysis

Default Rate By Delinquency in last 2 years & Number of Bankruptcy





- **For Bankruptcy, a High number of bankruptcies leads to default in every category of Delinquency in the last 2 years.**
- **For Delinquency in the last 2 years, applicants with delinquencies greater than or equal to 4 are having a high default rate for all of the bankrupt categories.**

The combination of delinquency above or equal to 4 with bankruptcy is an important combination for determining the default rate.

Conclusions

- A Higher Loan amount leads to a higher default rate. **This is one significant factor leading to Default.**
- Loans for a longer-term will have higher chances of defaulting. **This is one significant factor leading to Default.**
- Loans with high-interest rates will have higher chances of defaulting. **This is one significant factor leading to Default.**
- A lower grade means that the applicant is more likely to default. **This is one significant factor leading to Default.**
- Home Ownership as **"Others"** is more likely to default. **This is one significant factor leading to Default.**
- Lower the income, more is the chances to default i.e. **Applicant with Annual Income less than $40,000**. **This is one significant factor leading to Default.**
- **"Small businesses"** as purpose are generally risky and are most likely to default. **This is one significant factor leading to Default.**
- Applicants belonging to **State NV(Nevada), TN(Tenessee), and AK(Alaska)** are most likely to default.
- Higher the DTI, the higher the chances to default.
- Higher the Revolving Balance Utilization, the higher the chances of Default. **This is one significant factor leading to Default.**
- Applicants who have already been bankrupt are in very poor financial conditions and are more likely to Default. **This is one significant factor leading to Default.**
- High-Interest Rate with Longer Loan Term is risky and is an important combination for determining the default rate.
- Low grades such as **D, E, F, G with Bankruptcy** is an important combination for determining default rate.
- Home Ownership type as **"Other" and Low annual income** is an important combination for determining default rate.
- **Very High revolving balance utilization and "Small Business" as loan purpose** is an important combination for determining default rate.
- **Delinquency above and equal to 4 with bankruptcy** is an important combination for determining the default rate. **This is one significant combination leading to Default.**

Top Driving Factors Leading To Default

1. Higher Loan Amount
2. Loans for a longer-term
3. Loans with high-interest rates
4. Lower grade Applicants
5. Home Ownership as "Others"
6. Lower Annual Income Applicants
7. "Small businesses" as Loan Purpose
8. High Revolving Balance Utilization
9. Bankrupt Applicants
10. Number of Delinquencies above and equal to 4

Recommendations

1. If the applicant is having a low grade with a previous record of Bankruptcy then the loan should be disapproved.
2. If the requested Loan Amount is high with a longer loan repayment term then the applicant's background should be checked thoroughly before approving the loan.
3. If the Interest Rate on Loan Amount is high with a longer loan repayment term then the loan repayment term should be reduced.
4. Background verification of applicants with High Revolving Balance Utilization or a high number of delinquencies should be properly done before approving the loan.
5. When the purpose is "small business" then check the applicant thoroughly as they have a high tendency to default.
6. If the applicant has mentioned Home Ownership as "Others" and has a low annual income then the loan should be disapproved after a thorough check.